****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

R : Thu Apr 27 09:32:02 2000; MasPar time 41.60 Seconds
486.838 Million cell updates/sec

Tabular output not generated.

Title: >US-09-421-213-2
Description: (1-855) from US09421213B.pep
Perfect Score: 6382
Sequence: 1 MGSDRARKGGGGPKDFGAGL.....PGVYTRLPLFRDWIKENTGV 855

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 38.523; Variance 166.585; scale 0.231

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1808	28.3	241	1	W22987	Human serine protease	8.25e-161
2	768	12.0	798	1	R57283	Bovine enterokinase.	2.72e-59
3	677	10.6	418	1	R89435	Trypsin-like enzyme.	1.39e-50
4	663	10.4	232	1	R89430	Trypsin-like enzyme.	3.01e-49
5	618	9.7	416	1	W96812	A mouse serine proteas	5.72e-45
6	615	9.6	356	1	W46917	Amino acid sequence of	1.10e-44
7	600	9.4	297	1	W77304	Amino acid sequence of	2.91e-43
8	578	9.1	812	1	R83959	Complete mouse plasmin	3.52e-41
9	578	9.1	812	1	W94036	Murine plasminogen.	3.52e-41
10	578	9.1	812	1	W07585	Murine plasminogen seq	3.52e-41
11	567	8.9	271	1	W77302	Amino acid sequence of	3.85e-40
12	567	8.9	317	1	Y13391	Amino acid sequence of	3.85e-40
13	566	8.9	437	1	W51457	Human plasminogen frag	4.79e-40
14	566	8.9	546	1	R34427	Sequence of tissue pla	4.79e-40
15	566	8.9	790	1	R60519	Human 'Glu' plasminoge	4.79e-40
16	567	8.9	791	1	W34285	Human plasminogen.	3.85e-40
17	566	8.9	810	1	R08065	Human plasminogen from	4.79e-40
18	566	8.9	810	1	W31169	Plasminogen protein fo	4.79e-40
19	566	8.9	810	1	R34428	Sequence encoded by a	4.79e-40
20	565	8.9	811	1	R12933	Plasminogen mutein X1	5.95e-40
21	565	8.9	812	1	R12934	Plasminogen mutein X2	5.95e-40
22	565	8.9	813	1	R12935	Plasminogen mutein X3	5.95e-40
23	568	8.9	816	1	R12946	Plasminogen mutein T19	3.10e-40
24	561	8.8	233	1	W22986	Human serine protease	1.42e-39
25	560	8.8	461	1	W40284	Human Factor IX protea	1.76e-39
26	562	8.8	810	1	R12938	Plasminogen mutein T1	1.14e-39
27	562	8.8	811	1	R12943	Plasminogen mutein T13	1.14e-39
28	562	8.8	811	1	R12944	Plasminogen mutein T14	1.14e-39
29	562	8.8	811	1	R12945	Plasminogen mutein T17	1.14e-39
30	562	8.8	811	1	R12939	Plasminogen mutein T2	1.14e-39
31	562	8.8	812	1	R12948	Plasminogen muteins T2	1.14e-39
32	563	8.8	813	1	R12936	Plasminogen mutein X5	9.19e-40
33	561	8.8	813	1	R12937	Plasminogen mutein X6	1.42e-39
34	556	8.7	415	1	Y03203	Amino acid sequence of	4.21e-39
35	553	8.7	415	1	R35761	Factor IX (IX).	8.07e-39
36	555	8.7	454	1	R67710	Human Factor-IX.	5.23e-39
37	555	8.7	456	1	P40178	Part of the sequence o	5.23e-39
38	555	8.7	461	1	R05393	Mutant human factor IX	5.23e-39
39	555	8.7	461	1	P50311	Sequence of human fac	5.23e-39
40	555	8.7	461	1	P50302	Sequence of human fact	5.23e-39
41	557	8.7	810	1	R13221	Human plasminogen vari	3.39e-39
42	556	8.7	810	1	R13220	R561S human plasminoge	4.21e-39
43	556	8.7	810	1	R12406	R561E human plasminoge	4.21e-39
44	556	8.7	810	1	R13219	R561G human plasminoge	4.21e-39
45	558	8.7	816	1	R12942	Plasminogen mutein T8	2.72e-39

ALIGNMENTS

RESULT 1
ID W22987 standard; Protein; 241 AA.
AC W22987;
DT 08-OCT-1997 (first entry)
DE Human serine protease 67 (SP67).
KW Human; colon carcinoma; COLO 201; cell line; serine protease; SP67;
KW screening; inhibitor; treatment; disease.
OS Homo sapiens.
PN J09149790-A.
PD 10-JUN-1997.
PF 24-JUL-1996; 212196.
PR 29-SEP-1995; JP-275105.
PA (SUNR) SUNTORY LTD.
DR WPI; 97-357902/33.
DR N-PSDB; T79128.
PT Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67
PT - useful to screen for specific inhibitors, e.g. to search for, or
PT study agent for treatment of various diseases
PS Claim 1; Pages 12-13; 16pp; Japanese.
CC The present sequence is the human colon carcinoma COLO 201
CC cell line derived serine protease 67 (SP67), which can be used to
CC screen for specific inhibitors, e.g. to search for, or study an
CC agent for the treatment of various diseases.
SQ Sequence 241 AA;

Query Match 28.3%; Score 1808; DB 1; Length 241;
Best Local Similarity 99.6%; Pred. No. 8.25e-161;
Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Db      1 VVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    615 VVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTA 674

Db     61 FLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPD 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    675 FLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPD 734

Db    121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLPQQITPRMMCVGFL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    735 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLPQQITPRMMCVGFL 794

Db    181 SGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    795 SGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENTG 854

Db    241 V 241
          |
```

```
Qy    855 V 855


RESULT    2
ID    R57283 standard; Protein; 798 AA.
AC    R57283;
DT    08-MAR-1995  (first entry)
DE    Bovine enterokinase.
KW    Enterokinase; EK; heavy chain; light chain; catalytic domain;
KW    digestive disorder; cleavage; fusion protein; trypsinogen;
KW    trypsin; enzyme; PACE gene.
OS    Bos taurus.
FH    Key             Location/Qualifiers
FT    domain          1. .563
FT                    /label= heavy_chain_C-terminal
FT                    /note= "non-catalytic domain"
FT    domain          564. .798
FT                    /label= light_chain
FT                    /note= "catalytic domain"
PN    WO9416083-A.
PD    21-JUL-1994.
PI    13-JAN-1994; U00616.
PR    15-JAN-1993; US-005944.
PA    (GEMY ) GENETICS INST INC.
PI    Lavallie ER;
DR    WPI; 94-249229/30.
DR    N-PSDB; Q70104.
PT    New nucleic acid encoding enterokinase activity - and related
PT    vectors, host cells, expression products and antibodies are
PT    useful in treating digestive disorders and for cleaving fusion
PT    proteins
PS    Disclosure; Page 28-30; 50pp; English.
CC    The enterokinase (EK) (or the EK gene when used in gene therapy) is
CC    used to treat digestive disorders associated with low EK activity
CC    (esp. inability to process trypsinogen to trypsin).  For cleaving
CC    fusion proteins, recombinant EK catalytic domain is much more
CC    efficient than the native two-chain holoenzyme and is not
CC    contaminated by other proteolytic enzymes.  For expression of
CC    recombinant EK, the 1691-2398 DNA fragment was fused to the 3'-end of
CC    the signal peptide and pro-region of the human PACE gene.  The prod.
CC    could be expressed in CHO cells to produce a chimaeric prod. from
CC    which the pro-region as cleaved by endogenous PACE, providing mature
CC    EK catalytic domain.
SQ    Sequence   798 AA;

  Query Match           12.0%;  Score 768;  DB 1;  Length 798;
  Best Local Similarity  42.9%;  Pred. No. 2.72e-59;
  Matches   102;  Conservative   61;  Mismatches  68;  Indels   7;  Gaps   6;

Db    562 PKIVGGSDSREGAWPWVVALY-FDDQQVCGASLVSRDWLVSAAHC-VY--G-RNMEPSKW 616
          :::|||:|: || ||| |:|  : : ::|||||:| :||||||||  :   | |  :|: |
Qy    613 ARVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQW 672

Db    617 KAVLGLHMASNLTSPQIETRLIDQIVINPHYNKRRKNNDIAMMHLEMKVNYTDYIQPICL 676
           | ||||  |  ::| :: | : :|:: :| :|    : ||:|:: ||:  :|:  ::||||
Qy    673 TAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICL 732

Db    677 PEENQVFPPGRICSIAGWGALIYQGSTADVLQEADVPLLSNEKCQQQMPEYNITENMVCA 736
          |: ::|||:|:   ::|||   | |: | :|| ::: :::   |:  :|:  ||  |:|
Qy    733 PDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLPQ-QITPRMMCV 791

Db    737 GYEAGGVDSCQGDSGGPLMCQENNRWLL-AGVTSFGYQCALPNRPGVYARVPRFTEWI 793
          |: :|||||||||||||   | : ::: ||| | |  ||  |:||||:|:| | :||
Qy    792 GFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWI 849


RESULT    3
ID    R89435 standard; Protein; 418 AA.
AC    R89435;
DT    27-MAY-1996  (first entry)
DE    Trypsin-like enzyme.
KW    Trypsin; fibrinogen; thrombin; expectorant; respiratory disease;
KW    asthma; VIP; vasoactive intestinal peptide; influenza virus;
KW    protease; primer; PCR; amplification.
OS    Homo sapiens.
FH    Key             Location/Qualifiers
FT    peptide         1. .186
FT                    /label= sig_peptide
FT    protein         187. .418
FT                    /label= mat_protein
PN    AU9527248-A.
PD    08-FEB-1996.
PF    31-JUL-1995; 027248.
PR    29-JUL-1994; JP-178607.
PA    (TEIJ ) TEIJIN LTD.
PI    Masuda K,  Ogawa H,  Suga T,  Sugimoto Y,  Takagi K;
PI    Yamaoka K,  Yasuoka S;
DR    WPI; 96-117356/13.
DR    N-PSDB; T10689.
PT    Nucleic acid sequence encoding trypsin-like enzyme - which digests
PT    fibrinogen, used as expectorant in treatment of respiratory
PT    diseases, e.g. bronchial asthma
PS    Example 11; Page 47-49; 65pp; English.
CC    The overlap parts of 107 bp between the sequences given in
CC    T10698 and T10703 were identified, and thereby their identity
CC    was confirmed. It was confirmed from the sequence analysis,
CC    that these overlapping sequences contained a region encoding the
CC    amino acids of the N-terminus 20 residues of the trypsin-like
CC    enzyme isolated from the cough phlegm.  The sequences were
CC    ligated, and the desired trypsin-like enzyme gene cDNA sequence
CC    was determined (T10689).
SQ    Sequence   418 AA;

  Query Match           10.6%;  Score 677;  DB 1;  Length 418;
  Best Local Similarity  40.0%;  Pred. No. 1.39e-50;
  Matches   104;  Conservative   64;  Mismatches  76;  Indels  16;  Gaps  12;

Db    171 NECGAGPDLITLSEQRILGGTEAEEGSWPWQVSLR-LNNAHHCGGSLINNMWILTAAHCF 229
          ::|  |   :|  : |::|||:|:|| |||||||: |  :| ||:|||:  |:::||||:
Qy    600 KDCDCGLRSFT-RQARVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCY 658

Db    230 ---RS-N-SNPRDWIATSGI---ST-TFPKLR-MRVRNILIHNNYKSATHENDIALVRLE 279
             |:  |:| :| | |:  | : | :: |:: |: |  ::  | : ||||: ||
Qy    659 IDDRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELE 718

Db    280 NSVTFTKDIHSVCLPAATQNIPPGSTAYVTGWGAQEYAGHTVPELRQGQVRIISNDVCNA 339
          ::  ::  ::::||| |:: :|:| : |||||  :|:|     |: |::|:|:   |:
Qy    719 KPAEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCEN 778

Db    340 PHSYNGAILSGMLCAGVPQGGVDACQGDSGGPLVQ-EDSRRLWFIVGIVSWGDQCGLPDK 398
             :    | : |:| |   ||||:|||||||||  |   |: |  |:|||| |:  :|
Qy    779 LLP-Q-QITPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRI-FQAGVVSWGDGCAQRNK 835

Db    399 PGVYTRVTAYLDWIRQQTGI 418
          ||||||:  : |||:: ||:
Qy    836 PGVYTRLPLFRDWIKENTGV 855


RESULT    4
ID    R89430 standard; Protein; 232 AA.
AC    R89430;
DT    26-MAY-1996  (first entry)
DE    Trypsin-like enzyme.
KW    Trypsin; fibrinogen; thrombin; expectorant; respiratory disease;
KW    asthma; VIP; vasoactive intestinal peptide; influenza virus;
KW    protease.
OS    Homo sapiens.
FH    Key             Location/Qualifiers
FT    peptide         1. .20
FT                    /label= N-terminal
PN    AU9527248-A.
PD    08-FEB-1996.
PF    31-JUL-1995; 027248.
PR    29-JUL-1994; JP-178607.
PA    (TEIJ ) TEIJIN LTD.
PI    Masuda K,  Ogawa H,  Suga T,  Sugimoto Y,  Takagi K;
```

```
PI   Yamaoka K, Yasuoka S;
DR   WPI; 96-117356/13.
DR   N-PSDB; T10688.
PT   Nucleic acid sequence encoding trypsin-like enzyme - which digests
PT   fibrinogen, used as expectorant in treatment of respiratory
PT   diseases, e.g. bronchial asthma
PS   Claim 2; Page 56; 65pp; English.
CC   The nucleic acid may be used in the prodn. of the trypsin-like
CC   enzyme. The enzyme digests synthetic trypsin and thrombin
CC   substrates, fibrinogen (therefore used as an expectorant in the
CC   treatment of respiratory diseases, eg bronchial asthma) and
CC   vasoactive intestinal peptide, but does not digest IgA, IgG,
CC   albumin, alpha-1-antitrypsin or substance P.
CC   In constract to trypsin, the enzyme also inactivates influenza
CC   viruses, the Miyadera strain of NDV and the New Jersey strain
CC   of VZV.
SQ   Sequence   232 AA;

  Query Match           10.4%;  Score 663;  DB 1;  Length 232;
  Best Local Similarity  41.0%;  Pred. No. 3.01e-49;
  Matches  100;  Conservative   60;  Mismatches  69;  Indels  15;  Gaps  11;

Db       1 ILGGTEAEEGSWPWQVSLR-LNNAHHCGGSLINNMWILTAAHCF---RS-N-SNPRDWIA 54
           ::|||:|:|| |||||||: |  :| ||:|||:  |:::||||:   |:   |:| :| |
Qy     615 VVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTA 674

Db      55 TSGI---ST-TFPKLR-MRVRNILIHNNYKSATHENDIALVRLENSVTFTKDIHSVCLPA 109
             |:   |  : | ::  |:: |: |  ::  | : ||||: ||::  ::  ::::|||
Qy     675 FLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPD 734

Db     110 ATQNIPPGSTAYVTGWGAQEYAGHTVPELRQGQVRIISNDVCNAPHSYNGAILSGMLCAG 169
           |:: :|:| :  |||||  :|:|     |: |::|:|:   |:   :     | : |:| |
Qy     735 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLP-Q-QITPRMMCVG 792

Db     170 VPQGGVDACQGDSGGPLVQ-EDSRRLWFIVGIVSWGDQCGLPDKPGVYTRVTAYLDWIRQ 228
             ||||:|||||||||   |   |: |  |:|||||  |:  :|||||||:  : |||::
Qy     793 FLSGGVDSCQGDSGGPLSSVEADGRI-FQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKE 851

Db     229 QTGI 232
             ||:
Qy     852 NTGV 855


RESULT    5
ID   W96812 standard; Protein; 416 AA.
AC   W96812;
DT   21-APR-1999  (first entry)
DF   A mouse serine protease called hepsin.
F    ouse serine protease; hepsin; animal model; bone disease;
KW   one disorder; skeletal disorder; osteoporosis; Paget's disease;
KW   osteitis deformans; elevated bone alkaline phosphatase level.
OS   Mus musculus.
PN   WO9854307-A1.
PD   03-DEC-1998.
PF   29-MAY-1998; E03199.
PR   30-DEC-1997; US-000486.
PR   30-MAY-1997; US-866058.
PA   (SCHD ) SCHERING AG.
PA   (UNIW ) UNIV WASHINGTON.
PI   Sadler JE, Wu Q;
DR   WPI; 99-070213/06.
DR   N-PSDB; X15134.
PT   New nucleic acid functionally disrupts mouse hepsin gene  - used to
PT   provide transgenic mice with abnormally elevated blood alkaline
PT   phosphatase, useful as models for bone disorders
PS   Example 1; Fig 2; 29pp; English.
CC   The present sequence represents a mouse serine protease called hepsin.
CC   The specification describes a mammalian cell in which expression of a
CC   gene encoding hepsin has been functionally interrupted or suppressed.
CC   The products and methods provide an animal model for bone disease,
CC   and are useful to determine effective treatment for bone and skeletal
CC   disorders such as osteoporosis, Paget's disease and osteitis
CC   deformans, especially those associated with elevated bone alkaline
CC   phosphatase levels.
SQ   Sequence   416 AA;

  Query Match           9.7%;  Score 618;  DB 1;  Length 416;
  Best Local Similarity  39.5%;  Pred. No. 5.72e-45;
  Matches  102;  Conservative   55;  Mismatches  84;  Indels  17;  Gaps  16;

Db     149 CQDCGRRKLPVD-RIVGGQDSSLGRWPWQVSLRYDG-THLCGGSLLSGDWVLTAAHCF-P 205
           | ||| | :  : :|:||| |:  | |||||||:  |  |:||:||:| :|:::||||:
Qy     602 C-DCGLRSFTRQARVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

Db     206 ERN-R-V-LSRWRVFAGAVARTSPHAVQLGVQAVIY-HGGYLPFRDPTIDENSNDIALVH 261
           :|  |   ::|  | |  : | ::  |||     :    || :  :  :  ||||:
Qy     661 DRGFRYSDPTQWTAFLGLHDQ-SQRSAP-GVQERRLKRIISHPFFN-DFTFD-YDIALLE 716

Db     262 LSSSLPLTEYIQPVCLPAAGQALVDGKVCTVTGWGNTQFYGQQAMVLQEARVPIISNEVC 321
           |  :   :  ::|:||| |:: :  ||   |||||:||: |  |::|| : : :|:   |
Qy     717 LEKPAEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTC 776

Db     322 NSPDFYGNQIKPKMFCAGYPEGGIDACQGDSGGPFVCEDSISGTSRWRLCGIVSWGTGCA 381
           :  ::   |  |:| | |: ||:|:|||||||:   ::  |    :  |:|||| |||
Qy     777 E--NLLPQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEA-DG-RIFQA-GVVSWGDGCA 831

Db     382 LARKPGVYTKVTDFREWI 399
             ||||||::  ||:||
Qy     832 QRNKPGVYTRLPLFRDWI 849


RESULT    6
ID   W46917 standard; Peptide; 356 AA.
AC   W46917;
DT   02-JUL-1998  (first entry)
DE   Amino acid sequence of a novel human kallikrein.
KW   Kallikrein; HKLP; human; serine protease; drug screening; atagonist;
KW   agonist; treatment; hypertension; cardiac hypertrophy; arthritis;
KW   inflammatory disorder; blood clotting disorder.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 106
FT                   /note= "encoded by CRC"
FT   Misc_difference 168
FT                   /note= "encoded by YGG"
PN   WO9803665-A1.
PD   29-JAN-1998.
PF   21-JUL-1997; UK2724.
PR   22-JUL-1996; US-681151.
PA   (INCY-) INCYTE PHARM INC.
PI   Au-Young J, Bandman O, Braxton SM, Goli SK;
DR   WPI; 98-120785/11.
DR   N-PSDB; V16846.
PT   Human kallikrein polypeptide and DNA encoding it - useful for
PT   screening compounds useful in treatment of e.g. hypertension,
PT   cardiac hypertrophy and arthritis
PS   Claim 1; Fig 1A-E; 59pp; English.
CC   The present sequence represents a novel human kallikrein (HKLP).
CC   Kallikreins are a large family of homologous serine proteases that act in
CC   a variety of circulatory and immune system functions. The cDNA sequence
CC   encoding HKLP was isolated from a heart tissue library, and was
CC   identified in cDNA, Incyte clone 307474, through a computer generated
CC   search for amino acid sequence alignments. HKLP shows 35% identity to rat
CC   kallikrein and 33% identity to human kallikrein. Unlike rat kallikrein,
CC   HKLP is hydrophobic in the carboxy terminus and likely to remain
CC   membrane bound  The HKLP protein has 3 potential glycosylation sites.
CC   The cDNA sequence encoding HKLP, and vectors and host cells containing
CC   it are useful for the recombinant production of HKLP. HKLP is useful in
CC   drug screening for potential antagonists (or agonists). The HKLP protein
CC   and cDNA are useful in the treatment of conditions such as hypertension,
CC   cardiac hypertrophy, arthritis, inflammatory disorders and blood clotting
CC   disorders.
SQ   Sequence   356 AA;

  Query Match           9.6%;  Score 615;  DB 1;  Length 356;
  Best Local Similarity  37.3%;  Pred. No. 1.10e-44;
```

```
  Matches   104;  Conservative   61;  Mismatches  99;  Indels  15;  Gaps  13;

Db      79 LHEL-LVNGQS-CESRSKISLLCT-KQ-DCGXRPAARMNKRILGGRTSRPGRWPWQCSLQ 134
           |: | | :| : |:::   |     |: ||| |: :|   |::||  :  | |||| ||:
Qy     575 LNGLCLSKGNPECDGKEDCSDGSDEKDCDCGLRSFTRQA-RVVGGTDADEGEWPWQVSLH 633

Db     135 SEPSGHICGCVLIAKKWVLTVAHCF--E-G-R-ENAAVXKVVLGINNLDHPSV-FMQTRF 188
           :  |||||  ||: :|::: |||:  : | |  :::     ||:::  : |   :| |
Qy     634 ALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERR 693

Db     189 VKTIILHPRYSRAVVDYDISIVELSEDISETGYVRPVCLPNPEQWLEPDTYCYITGWGHM 248
           :| || || ::    ||||:::||     :: |||:|||:: : : :     :|||||
Qy     694 LKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFPAGKAIWVTGWGHT 753

Db     249 --GNKMPFKLQEGEVRIISLEHCQSYFDMKTITTRMICAGYESGTVDSCMGDSGGPLVCE 306
             |   :: || ||:|:|:   |:: :    || ||:| |: || |||| |||||||
Qy     754 QYGGTGALILQKGEIRVINQTTCENLLPQQ-ITPRMMCVGFLSGGVDSCQGDSGGPLSSV 812

Db     307 KPGGRWTLFGLTSWGSVCFSKVLGPGVYSNVSYFVEWIK 345
           : ||    |: |||  | :   ||||: :: | :|||
C      813 EADGRIFQAGVVSWGDGCAQRN-KPGVYTRLPLFRDWIK 850


RESULT    7
ID   W77304 standard; Protein; 297 AA.
AC   W77304;
DT   07-JAN-1999  (first entry)
DE   Amino acid sequence of SP003LA, a homologue of HELA2.
KW   Serine protease; regulation; cell activity; viability; HELA2; ATC2;
KW   BCOM3; testisin; fertility; suppressor; testicular germ cell cancer;
KW   seminoma; testis-specific expression; antitumour; sperm development;
KW   infertility; human; chromosome 16p13.3.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Disulfide_bond  1
FT                   /note= "likely to be involved in disulphide bonding"
FT   Disulfide_bond  36
FT                   /note= "likely to be involved in disulphide bonding"
FT   Misc_difference 51
FT                   /note= "likely to be a catalytic residue"
FT   Disulfide_bond  52
FT                   /note= "likely to be involved in disulphide bonding"
FT   Misc_difference 100
FT                   /note= "likely to be a catalytic residue"
FT   Disulfide_bond  134
FT                   /note= "likely to be involved in disulphide bonding"
FT   Disulfide_bond  167
FT                   /note= "likely to be involved in disulphide bonding"
FT   Disulfide_bond  190
FT                   /note= "likely to be involved in disulphide bonding"
FT   Disulfide_bond  201
FT                   /note= "likely to be involved in disulphide bonding"
FT   Misc_difference 205
FT                   /note= "likely to be a catalytic residue"
FT   Disulfide_bond  211
FT                   /note= "likely to be involved in disulphide bonding"
FT   Disulfide_bond  229
FT                   /note= "likely to be involved in disulphide bonding"
PN   WO9836054-A1.
PD   20-AUG-1998.
PF   13-FEB-1998; AU0085.
PR   18-NOV-1997; AU-000422.
PR   13-FEB-1997; AU-005101.
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
PI   Antalis TM, Hooper JD;
DR   WPI; 98-480768/41.
DR   N-PSDB; V59136.
PT   New serine protease(s) and kinase involved in regulating cell
PT   activity and viability - particularly the testis-specific protease
PT   HELA2 used for modulation of fertility and as tumour suppressor
PS   Example 15; Fig 20C; 167pp; English.
CC   W77302-04 represent HELA2 homologues. The genes are found in a cluster
CC   on chromosome 16p13.3. HELA2 was isolated from HeLa cells, and has
CC   homology to serine proteases. The protein is involved in or associated
CC   with regulation of cell activity and/or viability. Administration of
CC   recombinant HELA2 (also called testisin) is used to increase fertility.
CC   Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of
CC   testicular germ cell cancers (seminoma) and is also expressed in some
CC   non-testicular cancers (of colon, pancreas, prostate and ovary), so is
CC   a marker/potential therapeutic target for cancer. The promoter from the
CC   HELA2 gene is useful for testis-specific expression of other genes,
CC   e.g. for gene therapy or modulation of fertility. Drugs that block
CC   activity of HELA2 should have antitumour activity (other than in
CC   testis) while in testis recombinant HELA2 should stop growth of tumours
CC   and normalise sperm development (eliminating the need for orchidectomy).
CC   Identification of mutant forms of HELA2 can be used to diagnose
CC   infertility.
SQ   Sequence   297 AA;

  Query Match             9.4%;  Score 600;  DB 1;  Length 297;
  Best Local Similarity  38.1%;  Pred. No. 2.91e-43;
  Matches    98;  Conservative   70;  Mismatches  68;  Indels  21;  Gaps  14;

Db       1 CG-RP-RMLNRMVGGQDTQEGEWPWQVSIQRNG-SHFCGGSLIAEQWVLTAAHCF---RN 54
             || |:     |:||| |::||||||||||::  | :|:||:|||:  |:::||||:   |
Qy     604 CGLRSFTRQARVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRG 663

Db      55 -T-SETSLYQVLLGARQLVQPGPHAMYAR-VRQVESNPLYQGTASSADVALVELEAPVPF 111
             +|: :   :|| ::  | :: ::  | :::: |:|::   :   |:||:||| |  :
Qy     664 FRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEY 723

Db     112 TNYILPVCLPDPSVIFETGMNCWVTGWGSPSEEDLLPEPRILQKLAVPIIDTPKCNLLYS 171
           :: : |:||||:| :| :|:  ||||||  ::      : ||||  : :|:   |: | :
Qy     724 SSMVRPICLPDASHVFPAGKAIWVTGWG-HTQYGG-TGALILQKGEIRVINQTTCENLLP 781

Db     172 KDTEFGYQPKTIKNDMLCAGFEEGKKDACKGDSGGPLVCLVGQSWL-QAGVISWGEGCAR 230
             : ::   |:     |:| ||.|  |:| |||||||  : ::::: ||||:|||:|||:
Qy     782 -Q-QIT--PR-----MMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQ 832

Db     231 QNRPGVYIRVTAHHNWI 247
           :|:|||| |:   ::||
Qy     833 RNKPGVYTRLPLFRDWI 849


RESULT    8
ID   R83959 standard; protein; 812 AA.
AC   R83959;
DT   10-MAR-1996  (first entry)
DE   Complete mouse plasminogen molecule.
KW   Angiostatin; plasminogen; endothelial inhibitor; therapeutic;
KW   gene therapy.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   protein         1. .812
FT                   /note= "plasminogen"
FT   protein         98. .436
FT                   /note= "angiostatin"
PN   WO9529242-A1.
PD   02-NOV-1995.
PF   26-APR-1995; U05107.
PR   26-APR-1994; US-2[illegible]629.
PR   20-OCT-1994; US-326785.
PA   (CHIL-) CHILDRENS MEDICAL CENT.
PI   Cao Y,  Folkman MJ,  O'Reilly MS,  Sim KL;
DR   WPI; 95-382990/49.
PT   Endothelial inhibitor Angiostatin - useful to treat angiogenic
PT   mediated disease esp. angiogenesis and cancer.
PS   Disclosure; Fig 1; 108pp; English.
CC   Angiostatin (see R83960) is a plasminogen fragment starting at
CC   amino acid 98 of the complete plasminogen molecule.  Preferably,
CC   angiostatin has an amino acid sequence similar to that of the
CC   plasminogen fragment.  Angiostatin is an endothelial inhibitor,
CC   which reversibly inhibits proliferation of endothelial cells and
CC   thereby inhibits angiogenesis.  It is useful in the treatment of
CC   a human or animal with angiogenic mediated disease e.g. arthritis,
CC   macular degeneration, diabetic retinopathy or cancer.  Cells
```

```
CC   comprising angiostatin-coding sequences are useful for gene therapy
CC   of primary tumors.
SQ   Sequence   812 AA;

  Query Match           9.1%;  Score 578;  DB 1;  Length 812;
  Best Local Similarity  43.2%;  Pred. No. 3.52e-41;
  Matches   104;  Conservative   51;  Mismatches  67;  Indels  19;  Gaps  15;

Db     580 GRVVGGCVANPHSWPWQISLRTRFTGQHFCGGTLIAPEWVLTAAHC-LE-KSSR-PE-F- 634
           :|||||  |:   ||||:||::   | |:||::||:|:|:::|||| :: :: | ::
Qy     613 ARVVGGTDADEGEWPWQVSLHALGQG-HICGASLISPNWLVSAAHCYIDDRGFRYSDPTQ 671

Db     635 YKVILGAHEEYIRGL-DVQEISVAKLILEP--NN-R---DIALLKLSRPATITDKVIPAC 687
             :|| |::  |:   ||| : ::|  |  |:     ||||| | :||  : :| | |
Qy     672 WTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPIC 731

Db     688 LPSPNYMV-ADRTICYITGWGETQ-GTFGAGRLKEAQLPVIENKVCNRVEYLNNRVKSTE 745
           || :: :  | ::|  :|||| || |  ||  |  ::: ||:   |: :  |  :: :
Qy     732 LPDASHVFPAGKAI-WVTGWGHTQYGGTGALILQKGEIRVINQTTCENL--LPQQITPRM 788

.      746 LCAGQLAGGVDSCQGDSGGPLVCFEKDKYILQ-GVTSWGLGCARPNKPGVYVRVSRFVDW 804
           :| | |:|||||||||||||||   | |  |:| || ||| |||: |||||| |:: | ||
Qy     789 MCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDW 848

Db     805 I 805
           |
Qy     849 I 849


RESULT    9
ID   W94036 standard; protein; 812 AA.
AC   W94036;
DT   12-APR-1999  (first entry)
DE   Murine plasminogen.
KW   Plasminogen; kringle; endothelial; angiogenesis; tumour; leukaemia;
KW   rheumatoid arthritis; psoriasis; ocular angiogenic disease; ulcer;
KW   gene therapy; birth control; Crohn's disease; angiostatin.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   Domain          103. .181
FT                   /note= "Kringle 1"
FT   Domain          185. .262
FT                   /note= "Kringle 2"
FT   Domain          275. .352
FT                   /note= "Kringle 3"
FT   Domain          377. .454
FT                   /note= "Kringle 4"
FT   Domain          185. .352
FT                   /note= "Kringle 2-3"
FT   Domain          103. .352
FT                   /note= "Kringle 1-3"
FT   Domain          103. .262
FT                   /note= "Kringle 1-2"
FT   Domain          103. .454
FT                   /note= "Kringle 1-4"
FT   Domain          93. .470
FT                   /note= "Kringle 1-4BKLS"
PN   WO9854217-A1.
PD   03-DEC-1998.
PF   29-MAY-1998; U10979.
PR   30-MAY-1997; US-866735.
PA   (CHIL-) CHILDRENS MEDICAL CENT.
PI   Folkman MJ, OReilly MS;
DR   WPI; 99-059809/05.
PT   Use of plasminogen fragments - having an amino acid sequence similar
PT   to the kringle 1-5 region, for inhibiting endothelial cell
PT   proliferation and angiogenesis
PS   Examples; Fig 1A-B; 165pp; English.
CC   The invention relates to inhibition of endothelial cell proliferation
CC   that comprises administering to an endothelial cell a plasminogen
CC   fragment having an amino acid sequence similar to the kringle 1-5 region
CC   of a plasminogen molecule. The plasminogen fragments can be derived from
CC   murine, human, Rhesus, porcine or bovine plasminogens. The plasminogen
CC   fragments can be used for modulating angiogenesis and treating angiogenic
CC   -mediated disease e.g. solid tumours, blood born tumours (leukaemias),
CC   tumour metastasis, benign tumours (e.g. haemangiomas, acoustic neuromas,
CC   neurofibromas, trachomas and pyogenic granulomas), rheumatoid arthritis,
CC   psoriasis, ocular angiogenic diseases (e.g. diabetic retinopathy,
CC   retinopathy of prematurity, macular degeneration, corneal graft
CC   rejection, neovascular glaucoma, retrolental fibroplasia, rubeosis),
CC   Osler-Webber Syndrome, myocardial angiogenesis, haemophiliac joints,
CC   plaque neovascularisation, telangiectasia, angiofibroma, wound
CC   granulation. The fragments can also be used for treating excessive or
CC   abnormal stimulation of endothelial cells (intestinal adhesions), Crohn's
CC   disease, atherosclerosis, scleroderma and hypertrophic scars; as a birth
CC   control agent by preventing vascularisation required for embryo
CC   implantation; and in the treatment of diseases that have angiogenesis as
CC   a pathological consequence such as cat scratch disease (Rochele minalia
CC   quitosa) and ulcers (H. pylori). The nucleotide sequences encoding the
CC   plasminogen fragments can also be used for gene therapy. The products can
CC   be used for the production of antibodies and in detection and diagnosis.
CC   The present sequence represents a murine plasminogen sequence. The
CC   kringle regions that can be used in the invention are indicated in the
CC   features.
SQ   Sequence   812 AA;

  Query Match           9.1%;  Score 578;  DB 1;  Length 812;
  Best Local Similarity  43.2%;  Pred. No. 3.52e-41;
  Matches   104;  Conservative   51;  Mismatches  67;  Indels  19;  Gaps  15;

Db     580 GRVVGGCVANPHSWPWQISLRTRFTGQHFCGGTLIAPEWVLTAAHC-LE-KSSR-PE-F- 634
           :|||||  |:   ||||:||::   | |:||::||:|:|:::|||| :: :: | ::
Qy     613 ARVVGGTDADEGEWPWQVSLHALGQG-HICGASLISPNWLVSAAHCYIDDRGFRYSDPTQ 671

Db     635 YKVILGAHEEYIRGL-DVQEISVAKLILEP--NN-R---DIALLKLSRPATITDKVIPAC 687
             :|| |::  |:   ||| : ::|  |  |:     ||||| | :||  : :| | |
Qy     672 WTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPIC 731

Db     688 LPSPNYMV-ADRTICYITGWGETQ-GTFGAGRLKEAQLPVIENKVCNRVEYLNNRVKSTE 745
           || :: :  | ::|  :|||| || |  ||  |  ::: ||:   |: :  |  :: :
Qy     732 LPDASHVFPAGKAI-WVTGWGHTQYGGTGALILQKGEIRVINQTTCENL--LPQQITPRM 788

Db     746 LCAGQLAGGVDSCQGDSGGPLVCFEKDKYILQ-GVTSWGLGCARPNKPGVYVRVSRFVDW 804
           :| | |:|||||||||||||||   | |  |:| || ||| |||: |||||| |:: | ||
Qy     789 MCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDW 848

Db     805 I 805
           |
Qy     849 I 849


RESULT   10
ID   W07585 standard; protein; 812 AA.
AC   W07585;
DT   24-JUN-1997  (first entry)
DE   Murine plasminogen sequence.
KW   angiostatin; plasminogen; kringle; angiogenesis; cancer; arthritis;
KW   macular degeneration; diabetic retinopathy.
OS   Mus musculus.
PN   WO9635774-A2.
PD   14-NOV-1996.
PF   26-APR-1996; U05856.
PR   26-APR-1995; US-429743.
PR   22-FEB-1996; US-605598.
PR   08-MAR-1996; US-612788.
PA   (CHIL-) CHILDRENS MEDICAL CENT.
PI   Cao Y,  Folkman MJ,  Lin J,  Oreilly MS,  Sim KL;
DR   WPI; 96-518662/51.
PT   Use of angiostatin fragments or aggregates - for inhibiting
PT   endothelial cell proliferation and treating angiogenesis-mediated
PT   diseases, e.g. cancer, arthritis or diabetic retinopathy
PS   Disclosure; Fig 1; 203pp; English.
CC   The invention relates to new methods and compositions for
CC   inhibiting endothelial cell proliferation, using as active component
CC   an angiostatin fragment, a combination of angiostatin fragments, or
CC   aggregate angiostatin. The fragment is preferably derived from murine,
```

```
CC   human, Rhesus, porcine or bovine plasminogen and is a kringle 1,
CC   kringle 2, kringle 3, kringle 2-3, kringle 1-3, kringle 1-2, kringle
CC   1-4 or kringle 1-4BKLS protein. The aggregate angiostatin has a Mol. Wt.
CC   of 45-65 kD and is derived from a plasminogen fragment beginning at
CC   approximately amino acid number 98 of murine, human, Rhesus, porcine or
CC   bovine plasminogen. The active component can be used for treating
CC   angiogenesis-mediated diseases such as cancer, arthritis, macular
CC   degeneration and diabetic retinopathy. It can also be used to develop
CC   antibodies for use in diagnosis, detection and therapy.
CC   The present sequence is the whole murine plasminogen sequence from
CC   which fragments or aggregates may be derived for use in the invention.
SQ   Sequence   812 AA;

  Query Match            9.1%;  Score 578;  DB 1;  Length 812;
  Best Local Similarity  43.2%;  Pred. No. 3.52e-41;
  Matches   104;  Conservative   51;  Mismatches   67;  Indels   19;  Gaps   15;

Db     580 GRVVGGCVANPHSWPWQISLRTRFTGQHFCGGTLIAPEWVLTAAHC-LE-KSSR-PE-F- 634
           :|||||  |:   ||||:||::   | |:||::||:|:|:::|||| :: :: | ::
Qy     613 ARVVGGTDADEGEWPWQVSLHALGQG-HICGASLISPNWLVSAAHCYIDDRGFRYSDPTQ 671

Db     635 YKVILGAHEEYIRGL-DVQEISVAKLILEP--NN-R---DIALLKLSRPATITDKVIPAC 687
           :|| |::  |:   ||| : ::|  |  |:     ||||| | :||  : :| | |
Qy     672 WTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPIC 731

Db     688 LPSPNYMV-ADRTICYITGWGETQ-GTFGAGRLKEAQLPVIENKVCNRVEYLNNRVKSTE 745
           || :: :  | ::|  :|||| || |  ||  |  ::: ||:   |: :  |  :: :
Qy     732 LPDASHVFPAGKAI-WVTGWGHTQYGGTGALILQKGEIRVINQTTCENL--LPQQITPRM 788

Db     746 LCAGQLAGGVDSCQGDSGGPLVCFEKDKYILQ-GVTSWGLGCARPNKPGVYVRVSRFVDW 804
           :| | |:|||||||||||||  | |  |:| || ||| |||: ||||||  |:: | ||
Qy     789 MCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDW 848

Db     805 I 805
           |
Qy     849 I 849


RESULT   11
ID   W77302 standard; Protein; 271 AA.
AC   W77302;
DT   07-JAN-1999  (first entry)
DE   Amino acid sequence of SP001LA, a homologue of HELA2.
KW   Serine protease; regulation; cell activity; viability; HELA2; ATC2;
KW   BCOM3; testisin; fertility; suppressor; testicular germ cell cancer;
KW   seminoma; testis-specific expression; antitumour; sperm development;
KW   infertility; human; chromosome 16p13.3.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
     Disulfide_bond  29
FT                   /note= "likely to be involved in disulphide bonding"
FT   Misc_difference 44
FT                   /note= "likely to be a catalytic residue"
FT   Disulfide_bond  45
FT                   /note= "likely to be involved in disulphide bonding"
FT   Misc_difference 95
FT                   /note= "likely to be a catalytic residue"
FT   Disulfide_bond  115
FT                   /note= "likely to be involved in disulphide bonding"
FT   Disulfide_bond  129
FT                   /note= "likely to be involved in disulphide bonding"
FT   Misc_difference 160
FT                   /note= "encoded by GGAA"
FT   Disulfide_bond  162
FT                   /note= "likely to be involved in disulphide bonding"
FT   Disulfide_bond  181
FT                   /note= "likely to be involved in disulphide bonding"
FT   Disulfide_bond  192
FT                   /note= "likely to be involved in disulphide bonding"
FT   Misc_difference 196
FT                   /note= "likely to be a catalytic residue"
FT   Disulfide_bond  202
FT                   /note= "likely to be involved in disulphide bonding"
FT   Disulfide_bond  220
FT                   /note= "likely to be involved in disulphide bonding"
FT   Misc_difference 235
FT                   /note= "likely to be a catalytic residue"
PN   WO9836054-A1.
PD   20-AUG-1998.
PF   13-FEB-1998; AU0085.
PR   18-NOV-1997; AU-000422.
PR   13-FEB-1997; AU-005101.
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
PI   Antalis TM, Hooper JD;
DR   WPI; 98-480768/41..
DR   N-PSDB; V59134.
PT   New serine protease(s) and kinase involved in regulating cell
PT   activity and viability - particularly the testis-specific protease
PT   HELA2 used for modulation of fertility and as tumour suppressor
PS   Example 15; Fig 20A; 167pp; English.
CC   W77302-04 represent HELA2 homologues. The genes are found in a cluster
CC   on chromosome 16p13.3. HELA2 was isolated from HeLa cells, and has
CC   homology to serine proteases. The protein is involved in or associated
CC   with regulation of cell activity and/or viability. Administration of
CC   recombinant HELA2 (also called testisin) is used to increase fertility.
CC   Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of
CC   testicular germ cell cancers (seminoma) and is also expressed in some
CC   non-testicular cancers (of colon, pancreas, prostate and ovary), so is
CC   a marker/potential therapeutic target for cancer. The promoter from the
CC   HELA2 gene is useful for testis-specific expression of other genes,
CC   e.g. for gene therapy or modulation of fertility. Drugs that block
CC   activity of HELA2 should have antitumour activity (other than in
CC   testis) while in testis recombinant HELA2 should stop growth of tumours
CC   and normalise sperm development (eliminating the need for orchidectomy).
CC   Identification of mutant forms of HELA2 can be used to diagnose
CC   infertility.
SQ   Sequence   271 AA;

  Query Match            8.9%;  Score 567;  DB 1;  Length 271;
  Best Local Similarity  38.5%;  Pred. No. 3.85e-40;
  Matches    97;  Conservative   59;  Mismatches   78;  Indels   18;  Gaps   14;

Db       3 RVVGGEDSTDSEWPWIVSIQKNG-THHCAGSLLTSRWVITAAHCFKDN--L--NKPYLFS 57
           ||||| |: ::|||| ||::  |  | |::||::: |:::||||: |:  :  : |   :
Qy     614 RVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWT 673

Db      58 VLLGAWQLGNPGSRS-QKVGVAWVEPHPVYSWKEGACA-DIALVRLERSIQFSERVLPIC 115
           :||  : :  :: : |   : :: :|| :  :: :   ||||: ||:: ::|  | |||
Qy     674 AFLGLHDQSQRSAPGVQERRLKRIISHP-F-FNDFTFDYDIALLELEKPAEYSSMVRPIC 731

Db     116 LPDASIHLPPNTHCWISGWGSIQDGVPLPHPQTLQKLKVPIIDSEVCSHLYWRGAGQGPI 175
           |||||  :|:   |::|||  | |    :  |||  : :|:   | :|   : |  |
Qy     732 LPDASHVFPAGKAIWVTGWGHTQYGGT-G-ALILQKGEIRVINQTTCENLL---P-Q-QI 784

Db     176 TEDMLCAGYLEGERDACLGDSGGPLMC-QVDGAWLLAGIISWGEGCAERNRPGVYISLSA 234
           |  |:| |:| |  |:| ||||||   : ||  : ||::|||:|||:||:||||  |:
Qy     785 TPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPL 844

Db     235 HRSWVEKIVQGV 246
           | |: |   ||
Qy     845 FRDWI-KENTGV 855


RESULT   12
ID   Y13391 standard; Protein; 317 AA.
AC   Y13391;
DT   25-JUN-1999  (first entry)
DE   Amino acid sequence of protein PRO343.
KW   Secreted protein; transmembrane protein; human; enterocolitis;
KW   Zollinger-Ellison syndrome; gastrointestinal ulceration;
KW   congenital microvillus atrophy; skin disease; cell growth;
KW   abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
KW   Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
KW   fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
KW   anti-thrombotic; wound healing; tissue repair.
OS   Homo sapiens.
```

PN WO9914328-A2.
PD 25-MAR-1999.
PF 16-SEP-1998; U19330.
PR 25-NOV-1997; US-066840.
PR 17-SEP-1997; US-059113.
PR 17-SEP-1997; US-059115.
PR 17-SEP-1997; US-059117.
PR 17-SEP-1997; US-059119.
PR 17-SEP-1997; US-059121.
PR 17-SEP-1997; US-059122.
PR 17-SEP-1997; US-059184.
PR 18-SEP-1997; US-059263.
PR 18-SEP-1997; US-059266.
PR 15-OCT-1997; US-062125.
PR 17-OCT-1997; US-062285.
PR 17-OCT-1997; US-062287.
PR 21-OCT-1997; US-063486.
PR 24-OCT-1997; US-062814.
PR 24-OCT-1997; US-062816.
PR 24-OCT-1997; US-063045.
PR 24-OCT-1997; US-063120.
PR 24-OCT-1997; US-063121.
PR 24-OCT-1997; US-063127.
PR 24-OCT-1997; US-063128.
PR 27-OCT-1997; US-063329.
PR 27-OCT-1997; US-063327.
PR 28-OCT-1997; US-063541.
PR 28-OCT-1997; US-063542.
PR 28-OCT-1997; US-063544.
PR 28-OCT-1997; US-063549.
PR 28-OCT-1997; US-063550.
PR 28-OCT-1997; US-063564.
PR 29-OCT-1997; US-063435.
PR 29-OCT-1997; US-063704.
PR 29-OCT-1997; US-063732.
PR 29-OCT-1997; US-063738.
PR 29-OCT-1997; US-063734.
PR 29-OCT-1997; US-064215.
PR 29-OCT-1997; US-063735.
PR 31-OCT-1997; US-063870.
PR 31-OCT-1997; US-064103.
PR 03-NOV-1997; US-064248.
PR 07-NOV-1997; US-064809.
PR 12-NOV-1997; US-065186.
PR 17-NOV-1997; US-065846.
PR 18-NOV-1997; US-065693.
PR 21-NOV-1997; US-066120.
PR 21-NOV-1997; US-066364.
PR 24-NOV-1997; US-066772.
PR 24-NOV-1997; US-066466.
PR 24-NOV-1997; US-066770.
PR 24-NOV-1997; US-066511.
PR 24-NOV-1997; US-066453.
PA (GETH) GENENTECH INC.
PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
DR WPI; 99-229533/19.
DR N-PSDB; X52262.
PT New isolated human genes and polypeptides used in, e.g. treatment of
PT gastrointestinal ulceration
PS Claim 12; Fig 98; 320pp; English.
CC Y13344-403 represent secreted and transmembrane human proteins.
CC The cDNA sequences are obtained from cDNA libraries, prepared from
CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC The encoded polypeptides have specific uses based on their homology to
CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC associated with the preservation and maintenance of gastrointestinal
CC mucosa and the repair of acute and chronic mucosal lesions
CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC ulceration and congenital microvillus atrophy), skin diseases associated
CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC potent effects on cell growth and development, diseases related to growth
CC or survival of nerve cells including Parkinson's disease, Alzheimer's
CC disease, ALS, neuropathies or cancer. PRO265 can be used as for
CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
CC anti-thrombotic agent; PRO287 polypeptides and portions may have
CC therapeutic applications in wound healing and tissue repair; PRO317 can
CC be used for treating problems of the kidney, uterus, endometrium, blood
CC vessels, or related tissue, e.g. in the heart of genital tract.
SQ Sequence 317 AA;

Query Match 8.9%; Score 567; DB 1; Length 317;
Best Local Similarity 38.5%; Pred. No. 3.85e-40;
Matches 97; Conservative 59; Mismatches 78; Indels 18; Gaps 14;

```
Db     49 RVVGGEDSTDSEWPWIVSIQKNG-THHCAGSLLTSRWVITAAHCFKDN--L--NKPYLFS 103
           ||||| |: ::|||| ||::  |  | |::||::: |:::||||: |:  :  : |   :
Qy    614 RVVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWT 673

Db    104 VLLGAWQLGNPGSRS-QKVGVAWVEPHPVYSWKEGACA-DIALVRLERSIQFSERVLPIC 161
            :||  : :  :: : |   : :: :|| :  :: :   ||||: ||:: ::|  | |||
Qy    674 AFLGLHDQSQRSAPGVQERRLKRIISHP-F-FNDFTFDYDIALLELEKPAEYSSMVRPIC 731

Db    162 LPDASIHLPPNTHCWISGWGSIQDGVPLPHPQTLQKLKVPIIDSEVCSHLYWRGAGQGPI 221
          |||||  :|:    |::||| | |     :  |||  : :|:   | :|    : |  |
Qy    732 LPDASHVFPAGKAIWVTGWGHTQYGGT-G-ALILQKGEIRVINQTTCENLL---P-Q-QI 784

Db    222 TEDMLCAGYLEGERDACLGDSGGPLMC-QVDGAWLLAGIISWGEGCAERNRPGVYISLSA 280
          |  |:| |:| |  |:| ||||||   : ||  : ||::|||:|||:||:|||| |:
Qy    785 TPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPL 844

Db    281 HRSWVEKIVQGV 292
          | |: |   ||
Qy    845 FRDWI-KENTGV 855
```

RESULT 13
ID W51457 standard; protein; 437 AA.
AC W51457;
DT 02-SEP-1998 (first entry)
DE Human plasminogen fragment with neovascularisation inhibiting activity.
KW Human plasminogen; neovascularisation; angiostatin; inhibition;
KW elastase; Sepharose.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Disulfide_bond 4. .81
FT /label= Disulphide_bond
FT Disulfide_bond 25. .64
FT /label= Disulphide_bond
FT Disulfide_bond 53. .76
FT /label= Disulphide_bond
FT Disulfide_bond 108. .187
FT /label= Disulphide_bond
FT Disulfide_bond 129. .170
FT /label= Disulphide_bond
FT Disulfide_bond 158. .182
FT /label= Disulphide_bond
FT Disulfide_bond 194. .313
FT /label= Disulphide_bond
FT Disulfide_bond 204. .212
FT /label= Disulphide_bond
FT Disulfide_bond 234. .250
FT /label= Disulphide_bond
FT Disulfide_bond 326. .393
FT /label= Disulphide_bond
FT Disulfide_bond 356. .372
FT /label= Disulphide_bond
FT Disulfide_bond 383. .411
FT /label= Disulphide_bond
PN J10158300-A.
PD 16-JUN-1998.
PF 28-NOV-1996; 317250.
PR 28-NOV-1996; JP-317250.
PA (SUZM) SUZUKI KK.

```
DR   WPI; 98-393476/34.
PT   Human plasminogen derived polypeptide - has neovascularisation
PT   inhibiting activity
PS   Claim 1; Page 2; 16pp; Japanese.
CC   The invention relates to a neovascularisation inhibitor which comprises
CC   amino acids 355-791 of human plasminogen. Also claimed are a method for
CC   the preparation of angiostatin, and angiostatin prepared by this method.
CC   The human plasminogen protein fragment is prepared by: (a) applying human
CC   plasminogen to a lysine Sepharose column to separate it into plasminogen
CC   form 1 and form 2; (b) separating plasminogen form 1 and form 2 and
CC   digesting them with elastase; (c) fractionating the elastase-decomposed
CC   product of form 1 plasminogen and form 2 plasminogen in a lysine
CC   Sepharose column; (d) collecting the fractions bound to the lysine
CC   Sepharose column; (e) further fractionating the form 2 plasminogen using
CC   an Aminohexal Sepharose column; and (f) collecting the fraction bound to
CC   the Aminohexal Sepharose column. This human plasminogen fragment can be
CC   used to inhibit growth of vascular endothelial cells. The present
CC   sequence represents amino acids 355-791 of human plasminogen.
SQ   Sequence   437 AA;

     ry Match            8.9%;  Score 566;  DB 1;  Length 437;
     t Local Similarity  43.3%;  Pred. No. 4.79e-40;
  Matches   104;  Conservative   47;  Mismatches  71;  Indels  18;  Gaps  13;

Db     206 GRVVGGCVAHPHSWPWQVSLRTRFGM-HFCGGTLISPEWVLTAAHC-LE-KSPR---PSS 259
           :|||||  |    |||||||:: :|  |:||::||||:|:::|||| :: :: |   |:
Qy     613 ARVVGGTDADEGEWPWQVSLHA-LGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQ 671

Db     260 YKVILGAHQEVNL-EPHVQE--IE--VSRLFLEP-TRK-DIALLKLSSPAVITDKVIPAC 312
           :|| |::     | |||  :   :|: |::  |   ||||| |  ||  : :| | |
Qy     672 WTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPIC 731

Db     313 LPSPNYVVADRTECFITGWGETQ-GTFGAGLLKEAQLPVIENKVCNRYEFLNGRVQSTEL 371
           || :: | :      :|||| || |  || :|  ::: ||:   |:  ::|  :: :  :
Qy     732 LPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE--NLLPQQITPRMM 789

Db     372 CAGHLAGGTDSCQGDSGGPLVCFEKDKYILQ-GVTSWGLGCARPNKPGVYVRVSRFVTWI 430
           | | |:|| ||||||||||   | | |:| || ||| |||: |||||| |:: |  ||
Qy     790 CVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWI 849


RESULT   14
ID   R34427 standard; Protein; 546 AA.
AC   R34427;
DT   17-AUG-1993  (first entry)
DE   Sequence of tissue plasminogen activator (t-PA)/plasminogen
DE   hybrid protein.
KW   Zymogen; fibrinolytic activity; cleavage.
O    Synthetic.
I    JS5200340-A.
PL   06-APR-1993.
PF   22-MAY-1987; 053412.
PR   22-MAY-1987; US-053412.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Foster DC,  Mulvihill ER,  Ohara PJ,  Pingel K,  Yoshitake S;
DR   WPI; 93-133739/16.
DR   N-PSDB; Q40318.
PT   Human tissue plasminogen activator single chain form fibrinolytic
PT   agent - comprises thrombin cleavable zymogen stimulating amido
PT   lytic activity, for lysing clots in heart attack and stroke
PT   victims and suppressing fibrin matrix
PS   Example; Fig 8A, 8B, 8C; 22pp; English.
CC   A hybrid DNA sequence was constructed which encoded a protein
CC   consisting of the entire amino-terminal portion of t-PA (up to
CC   the cysteine at posn. 261) joined to the serine protease domain
CC   of plasminogen beginning at amino acid 541 (just to the amino-
CC   terminal side of the normal activation site). This hybrid protein
CC   was designated "PAP".
SQ   Sequence   546 AA;

  Query Match            8.9%;  Score 566;  DB 1;  Length 546;
  Best Local Similarity  43.3%;  Pred. No. 4.79e-40;
  Matches   104;  Conservative   47;  Mismatches  71;  Indels  18;  Gaps  13;

Db     315 GRVVGGCVAHPHSWPWQVSLRTRFGM-HFCGGTLISPEWVLTAAHC-LE-KSPR---PSS 368
           :|||||  |    |||||||:: :|  |:||::||||:|:::|||| :: :: |   |:
Qy     613 ARVVGGTDADEGEWPWQVSLHA-LGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQ 671

Db     369 YKVILGAHQEVNL-EPHVQE--IE--VSRLFLEP-TRK-DIALLKLSSPAVITDKVIPAC 421
           :|| |::     | |||  :   :|: |::  |   ||||| |  ||  : :| | |
Qy     672 WTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPIC 731

Db     422 LPSPNYVVADRTECFITGWGETQ-GTFGAGLLKEAQLPVIENKVCNRYEFLNGRVQSTEL 480
           || :: | :      :|||| || |  || :|  ::: ||:   |:  ::|  :: :  :
Qy     732 LPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE--NLLPQQITPRMM 789

Db     481 CAGHLAGGTDSCQGDSGGPLVCFEKDKYILQ-GVTSWGLGCARPNKPGVYVRVSRFVTWI 539
           | | |:|| ||||||||||   | |  |:| || ||| |||: |||||| |:: |  ||
Qy     790 CVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWI 849


RESULT   15
ID   R60519 standard; Protein; 790 AA.
AC   R60519;
DT   22-MAR-1995  (first entry)
DE   Human 'Glu' plasminogen.
KW   Serine protease; Factor-Xa; recognition site; plasminogen; kringle;
KW   fusion protein cleavage; protein folding; primer;
KW   polymerase chain reaction; amplification.
OS   Homo sapiens.
PN   WO9418227-A.
PD   18-AUG-1994.
PF   04-FEB-1994; DK0054.
PR   04-FEB-1993; DK-000130.
PR   05-FEB-1993; DK-000139.
PR   03-DEC-1993; WO-G02492.
PA   (DENZ-) DENZYME APS.
PI   Etzerodt M,  Holtet TL,  Thogersen HC;
DR   WPI; 94-279681/34.
PT   Refolding of polypeptide molecules - using a cyclic process
PT   involving denaturing and renaturing conditions to produce a
PT   correctly folded prod
PS   Disclosure; Page 148-50; 202pp; English.
CC   cDNA encoding kringle domains 1 and 4 of human plasminogen (full
CC   sequence given in R60519) was PCR amplified using primers given in
CC   Q71268-71.  Amplified cDNA was linked to a sequence encoding the
CC   Factor-Xa cleavage site (given in R60503), subcloned in vector
CC   pLcIIMLCH6 so that it was linked to a hexahistidine-encoding
CC   sequence and expressed in E. coli QY13.  The fusion protein was
CC   purified on an Ni2+-activated NTA-agarose column.  A cyclic
CC   procedure was used to obtain correctly folded recombinant protein.
SQ   Sequence   790 AA;

  Query Match            8.9%;  Score 566;  DB 1;  Length 790;
  Best Local Similarity  43.3%;  Pred. No. 4.79e-40;
  Matches   104;  Conservative   47;  Mismatches  71;  Indels  18;  Gaps  13;

Db     559 GRVVGGCVAHPHSWPWQVSLRTRFGM-HFCGGTLISPEWVLTAAHC-LE-KSPR---PSS 612
           :|||||  |    |||||||:: :|  |:||::||||:|:::|||| :: :: |   |:
Qy     613 ARVVGGTDADEGEWPWQVSLHA-LGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQ 671

Db     613 YKVILGAHQEVNL-EPHVQE--IE--VSRLFLEP-TRK-DIALLKLSSPAVITDKVIPAC 665
           :|| |::     | |||  :   :|: |::  |   ||||| |  ||  : :| | |
Qy     672 WTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPIC 731

Db     666 LPSPNYVVADRTECFITGWGETQ-GTFGAGLLKEAQLPVIENKVCNRYEFLNGRVQSTEL 724
           || :: | :      :|||| || |  || :|  ::: ||:   |:  ::|  :: :  :
Qy     732 LPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE--NLLPQQITPRMM 789

Db     725 CAGHLAGGTDSCQGDSGGPLVCFEKDKYILQ-GVTSWGLGCARPNKPGVYVRVSRFVTWI 783
           | | |:|| ||||||||||   | |  |:| || ||| |||: |||||| |:: |  ||
Qy     790 CVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWI 849


Search completed: Thu Apr 27 09:33:29 2000
Job time : 87 secs.
```

```
44     556
45     558                                    Plasminogen mutein T8    2.[illegible]
```

ALIGNMENTS

```
RESULT    1
ID   W22987 standard; Protein; 241 AA.
AC   W22987;
DT   08-OCT-1997  (first entry)
DE   Human serine protease 67 (SP67).
KW   Human; colon carcinoma; COLO 201; cell line; serine protease; SP67;
KW   screening; inhibitor; treatment; disease.
OS   Homo sapiens.
PN   J09149790-A.
PD   10-JUN-1997.
PF   24-JUL-1996; 212196.
PR   29-SEP-1995; JP-275105.
PA   (SUNR ) SUNTORY LTD.
DR   WPI; 97-357902/33.
DR   N-PSDB; T79128.
PT   Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67
PT   - useful to screen for specific inhibitors, e.g. to search for, or
PT   study agent for treatment of various diseases
PS   Claim 1; Pages 12-13; 16pp; Japanese.
CC   The present sequence is the human colon carcinoma COLO 201
CC   cell line derived serine protease 67 (SP67), which can be used to
CC   screen for specific inhibitors, e.g. to search for, or study an
CC   agent for the treatment of various diseases.
SQ   Sequence   241 AA;

  Query Match             28.3%;  Score 1808;  DB 1;  Length 241;
  Best Local Similarity   99.6%;  Pred. No. 8.25e-161;
  Matches  240;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Db        1 VVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTV 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      615 VVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTA 674

Db       61 FLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPD 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      675 FLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPD 734

Db      121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLPQQITPRMMCVGFL 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      735 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLPQQITPRMMCVGFL 794

Db      181 SGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENTG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      795 SGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENTG 854

Db      241 V 241
            |
```